Robust Optimization in Simulation

Dellino, G.; Kleijnen, Jack P.C.; Meloni, C.

Publication date:
2009

CentER

Discussion Paper

No. 2009–82

ROBUST OPTIMIZATION IN SIMULATION: TAGUCHI AND KRIGE COMBINED

By Gabriella Dellino, Jack P.C. Kleijnen, Carlo Meloni

October 2009

ISSN 0924-7815

Robust optimization in simulation: Taguchi and Krige combined

Gabriella Dellino
Department of Information Engineering, University of Siena,
Via Roma 56, 53100 Siena, Italy, {dellino@dii.unisi.it}

Jack P.C. Kleijnen
Department of Information Management / CentER,
Tilburg School of Economics and Management, Tilburg University, Postbox 90153, 5000 LE Tilburg, Netherlands, kleijnen@uvt.nl

Carlo Meloni
Department of Electrical Engineering and Electronics, Polytechnic of Bari,
Via E. Orabona 4, 70125 Bari, Italy, meloni@deemail.poliba.it

Optimization of simulated systems is the goal of many methods, but most methods assume known environments. We, however, develop a 'robust' methodology that accounts for uncertain environments. Our methodology uses Taguchi's view of the uncertain world, but replaces his statistical techniques by Kriging. We illustrate the resulting methodology through classic Economic Order Quantity (EOQ) inventory models. Our results suggest that robust optimization requires order quantities that differ from the classic EOQ. We also compare our latest results with our previous results that do not use Kriging but Response Surface Methodology (RSM).

Key words: Statistics, Design of experiments; Inventory-Production, Simulation; Decision analysis: Risk

JEL: C0, C1, C9

---

1. Introduction

In practice, some inputs of a given simulation model are uncertain so the optimum solution that is derived—ignoring these uncertainties—may be wrong. Strategic decision-making in such an uncertain world may use *Taguchi*'s approach, originally developed to help Toyota design 'robust' cars; i.e., cars that perform reasonably well in many circumstances; see Beyer and Sendhoff (2007), Kleijnen (2008), Park et al. (2006), Taguchi (1987) and Wu and Hamada (2000).

We use Taguchi's view of the world, but not his statistical methods. These methods use rather restrictive designs and analysis models, so we use *Kriging*. This Kriging applied to simulation or 'computer' models is explained by Sacks et al. (1989) and Santner et al. (2003). Kriging gives metamodels—also called response surfaces, surrogates, emulators, auxiliary models, repromodels, etc.; see Barton and Meckesheimer (2006) (we give so many synonyms because simulation is used in many disciplines, with their own terminologies). These metamodels run much faster than the underlying—possibly computationally expensive—simulation models; e.g., it takes 32 hours per run, on a high performance computer in the aerospace-engineering case study reported by Oberguggenberger et al. (2009). Kriging treats the simulation model as a black box; i.e., only the Input/Output (I/O) of the simulation model are observed. (Black-box methods have wider applicability but lower efficiency than white-box methods such as perturbation analysis and the score function.)

Moreover, we combine Kriging metamodeling with *Non Linear Programming* (NLP). In our NLP approach we select one of the multiple simulation outputs as the goal or objective, while the remaining outputs must satisfy given constraints (thresholds).

This combination of Kriging and NLP gives an estimate of the robust solution of the simulation optimization problem. Finally, in the NLP model we change specific threshold values for the constrained simulation outputs, to estimate the *Pareto frontier*.

We also compare our results with the results of a previous article in which we used Response Surface Methodology (RSM) as a heuristic to optimize the simulated system; see Dellino et al. (2009). We use Matlab software for the various components of our heuristic, because this software is well documented and is used by many simulationists.

Note that our methodology may also be applied to study so-called implementation errors, which are studied by Stinstra and den Hertog (2008).

The rest of this article is organized as follows. Section 2 summarizes Taguchi's worldview. Section 3 summarizes RSM for robust optimization. Section 4 summarizes Kriging (so readers familiar with Kriging may skip this text) and the use of Kriging for robust optimization. Section 5 illustrates the new methodology through the classic EOQ simulation model (which has a known I/O function and is a building block for more complicated and realistic supply-chain simulation models). Section 6 presents our conclusions and possible topics for future research.

2. Taguchi's worldview

Taguchi (1987) distinguishes between two types of factors (inputs, variables): (i) *decision* or *control* factors, which we denote by d_j ($j = 1, \dots, k$), and (ii) *environmental* or *noise* factors (say) e_g ($g = 1, \dots, c$). He assumes a single output (say) w. Taguchians focus on the mean and the variance of this output. By definition, the decision factors are under the control of the users; e.g., in inventory management, the order quantity is supposed to be controllable. The environmental factors are not controlled by the users; e.g., the demand rate in inventory management is a noise factor.

We adopt Taguchi's view, but not his statistical methods—which have been criticized by many statisticians; see Nair (1992). Instead of these methods we use Kriging including designs such as Latin Hypercube Sampling (LHS). Our reason for selecting Kriging is that the experimental area in simulation experiments may be much larger than in physical experiments, so a low-order polynomial may be an inadequate approximation (nonvalid metamodel). Our main reason for choosing a non-Taguchian design is that simulation experiments enable the exploration of many more factors, factor levels, and combinations of factor levels than real-life (physical) experiments do. For a further discussion of various metamodels and designs in simulation we refer to Kleijnen (2008) and Kleijnen et al. (2005).

Moreover, we do not use a *scalar* Taguchian loss function such as the signal-to-noise or mean-to-variance ratio; instead we allow the output to have a statistical distribution that we characterize through its mean and standard deviation. We formulate a NLP problem in which one of these characteristics (e.g., the mean of the primary simulation output) is the goal function to be minimized, while the other characteristics (e.g., the standard deviation of the goal output) must meet given constraints (Lehman et al. (2004) also minimize the mean while satisfying a constraint on the variance; they use a Bayesian approach). Next we change the thresholds (right-hand sides) in these constraints, and find the Pareto-optimal efficiency frontier—briefly called the *Pareto frontier*. Also see Beyer and Sendhoff (2007), Myers and Montgomery (1995, p. 491), Park et al. (2006), Wu et al. (2008).

3. RSM and robust optimization

In this section we summarize Dellino et al. (2009), who use RSM for robust simulation-optimization; this approach guides our Kriging approach. Like Myers and Montgomery

(1995), Dellino et al. use the following RSM metamodel:

$$y = \beta_0 + \boldsymbol{\beta}'\mathbf{d} + \mathbf{d}'\mathbf{B}\mathbf{d} + \boldsymbol{\gamma}'\mathbf{e} + \mathbf{d}'\boldsymbol{\Delta}\mathbf{e} + \epsilon \tag{1}$$

where y denotes the regression predictor of the expected (mean) simulation output $E(w)$, ϵ the residual with $E(\epsilon) = 0$ if this metamodel has no lack-of-fit (i.e., this metamodel is valid) and with constant variance σ_ϵ^2, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$ the first-order effects of the control variables $\mathbf{d} = (d_1, \ldots, d_k)'$, $\mathbf{B}$ the $k \times k$ symmetric matrix with the purely quadratic effects $\beta_{j;j}$ on the main diagonal and half the interaction effects $\beta_{j;j'}/2$ off the diagonal, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_c)'$ the first-order effects of the noise factors $\mathbf{e} = (e_1, \ldots, e_c)'$, and $\boldsymbol{\Delta} = (\delta_{j;g})$ the 'control-by-noise' two-factor interactions.

To examine whether the assumed metamodel (1) is an *adequate approximation*, Dellino et al. use leave-one-out cross-validation which eliminates one I/O combination, recomputes the estimates, and compares the estimated output with the output eliminated.

Under the assumption $E(\mathbf{e}) = \boldsymbol{\mu}_\mathbf{e}$ it is easy to derive that (1) implies

$$E(y) = \beta_0 + \boldsymbol{\beta}'\mathbf{d} + \mathbf{d}'\mathbf{B}\mathbf{d} + \boldsymbol{\gamma}'\boldsymbol{\mu}_\mathbf{e} + \mathbf{d}'\boldsymbol{\Delta}\boldsymbol{\mu}_\mathbf{e}\,. \tag{2}$$

Under the assumption $\mathbf{cov}(\mathbf{e}) = \boldsymbol{\Omega}_\mathbf{e}$ (1) implies

$$var(y) = (\boldsymbol{\gamma}' + \mathbf{d}'\boldsymbol{\Delta})\boldsymbol{\Omega}_\mathbf{e}(\boldsymbol{\gamma} + \boldsymbol{\Delta}'\mathbf{d}) + \sigma_\epsilon^2 = \mathbf{l}'\boldsymbol{\Omega}_\mathbf{e}\mathbf{l} + \sigma_\epsilon^2. \tag{3}$$

The mean and variance in (2) and (3) may be estimated through plugging in the Ordinary Least Squares (OLS) estimates of the regression coefficients in (1) into the right-hand sides of (2) and (3). (Del Castillo (2007, pp. 250–253) shows how the resulting bias in the estimated variance might be eliminated.)

Dellino et al. minimize the resulting estimated mean $\widehat{y}$, while keeping the estimated standard deviation $\widehat{\sigma_y}$ below a given Threshold T; they use the standard deviation instead of the variance because the standard deviation has the same scale as the mean. They solve this constrained minimization problem that is nonlinear in the decision variables $\mathbf{d}$; this gives the estimated robust decision variables $\widehat{\mathbf{d}^+}$. Next they vary T, which may give different solutions $\widehat{\mathbf{d}^+}$ with corresponding $\widehat{y}$ and $\widehat{\sigma_y}$. They then collect the pairs $(\widehat{y}, \widehat{\sigma_y})$ to estimate the Pareto frontier. Finally, they estimate the variability of this frontier through parametric bootstrapping of the OLS estimates that gave $\widehat{y}$ and $\widehat{\sigma_y}$ (we shall return to bootstrapping, in Section 4.2. They illustrate their methodology through the same EOQ model that we shall use in Section 5.

4. Kriging and robust optimization

In this section, we first summarize the basics of Kriging; next we discuss how to use Kriging for robust optimization.

4.1. Kriging basics

We base this subsection on Kleijnen (2008) and Kleijnen (2009). Typically, Kriging models are fitted to data that are obtained for larger, global experimental areas than the small, local areas used in low-order polynomial regression metamodels such as (1). Originally, Kriging was developed in geostatistics—also known as spatial statistics—by the South African mining engineer Danie Krige; see the classic geostatistics textbook Cressie (1993). Later on, Kriging was often applied to the I/O data of deterministic simulation models; see the classic article Sacks et al. (1989). Recently, Kriging has also been applied to random (stochastic) simulation; see Ankenman et al. (2009).

Kriging uses the *linear* predictor

$$y = \boldsymbol{\lambda}'\mathbf{w} \tag{4}$$

where—unlike the regression coefficients in (1)—the weights $\boldsymbol{\lambda}$ are not constants but decrease with the *distance* between the 'new' input combination to be predicted and the 'old' input combinations that have already been simulated and resulted in the simulation outputs $\mathbf{w}$.

Our heuristic uses the simplest type of Kriging called *Ordinary* Kriging, which assumes

$$w = \mu + \delta \tag{5}$$

where w is the simulation output (which depends on the input combination), μ is the simulation output averaged over the whole experimental area, and δ is the additive noise that forms a stationary covariance process (so its covariances decrease with the distances or 'lags' between the simulation input combinations) with zero mean. Note that Ankenman et al. (2009) call δ the 'extrinsic noise'; they add another term to (5) and call it 'intrinsic noise', which is caused by the Pseudo-Random Numbers (PRN) that are used in random simulation so the same input combination still shows intrinsic noise (this intrinsic noise has variance that may vary with the input combination, and is correlated if common PRN are used to simulate various input combinations).

If (5) holds, then the *optimal* weights in (4) can be proven to be

$$\lambda_o = \boldsymbol{\Gamma}^{-1}\left[\boldsymbol{\gamma} + \mathbf{1}\frac{1 - \mathbf{1}'\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}}{\mathbf{1}'\boldsymbol{\Gamma}^{-1}\mathbf{1}}\right] \tag{6}$$

where $\mathbf{\Gamma} =(cov(w_i, w_{i'}))$ with $i, i' = 1, \ldots, n$ is the $n \times n$ matrix with the covariances between the n old outputs, and $\boldsymbol{\gamma} = (cov(w_i, w_0))$ is the n-dimensional vector with the covariances between the old outputs w_i and w_0, the output of the combination to be predicted which may be either new or old. Obsviously $\boldsymbol{\gamma}$ varies with w_0, so λ_o in (6) varies with w_0.

The covariances $\mathbf{\Gamma}$ and $\boldsymbol{\gamma}$ in (6) are often based on the *Gaussian correlation function*

$$\exp\left(-\sum_{j=1}^{k} \theta_j h_j^2\right) = \prod_{j=1}^{k} \exp\left(-\theta_j h_j^2\right) \tag{7}$$

where h_j denotes the distance between input j of the new and the old combinations, and θ_j denotes the importance of input j (the higher θ_j is, the less effect input j has).

Substituting the correlation function (7) into (6) implies that the weights are relatively high for inputs close to the input to be predicted. Moreover, the optimal weights (6) imply that for an old input the predictor equals the observed simulation output at that input (all weights are zero except the weight of the observed output); i.e., the Kriging predictor is an *exact interpolator*. (RSM uses OLS, which minimizes the Sum of Squared Residuals so it is not an exact interpolator; Kriging accounting for the intrinsic noise in random simulation is not an exact interpolator either.)

So, the optimal weights (6) depend on the correlation function (7)—but this correlation function is unknown. Consequently, the parameter values θ_j in (7) must be estimated. The standard Kriging software and literature uses Maximum Likelihood Estimators (MLEs) assuming the noise δ in (5) is (multivariate) Normally (Gaussian) distributed. We estimate the correlation functions and the corresponding optimal weights through DACE, which is free-of-charge Matlab software that is well documented by Lophaven et al. (2002).

To get the I/O simulation data to which the Kriging model is fitted, simulation analysts often use LHS. This LHS assumes that a valid metamodel is more complicated than a low-order polynomial (which is assumed in RSM). LHS does not assume a specific metamodel or I/O function. Instead, LHS tries to fill the design space formed by the simulation inputs; i.e., LHS is a space-filling design (references and websites for other space-filling designs are given by Kleijnen (2008, pp. 127–130)). We shall further discuss LHS in our EOQ example in Section 5.

4.2. Two Kriging approaches to robust simulation-optimization

To solve robust simulation-optimization problems, we propose the following two approaches using Kriging metamodels:

- Inspired by Dellino et al. (2009), we fit *two* Kriging metamodels, namely one model for the mean and one for the standard deviation—both estimated from the *simulation*'s I/O data.

- Inspired by Lee and Park (2006), we fit a *single* Kriging metamodel to a relatively small number (say) n of combinations of the decision variables $\mathbf{d}$ and the environmental variables $\mathbf{e}$. Next, we use this metamodel to compute the *Kriging predictions* for the simulation output w for $N \gg n$ combinations of $\mathbf{d}$ and $\mathbf{e}$ accounting for the distribution of $\mathbf{e}$.

In the first approach, we select the input combinations for the simulation model through a *crossed* (combined) design for the decision and environmental factors (as is usual in Taguchian design); i.e., we combine the (say) n_d combinations of the decision variables $\mathbf{d}$ with the n_e combinations of the environmental variables $\mathbf{e}$. These n_d combinations are *space-filling*, so we can avoid extrapolation when using the Kriging metamodels to obtain predictions; Kriging is known to give bad extrapolators. The n_e combinations are *sampled* from their input distribution; we use LHS for this sampling. Simulating these $n_d \times n_e$ combinations gives the outputs $w_{i,j}$ with $i = 1, \ldots, n_d$ and $j = 1, \ldots, n_e$. These I/O data enable the computation of the following estimators of the n_d conditional means and variances:

$$\overline{w_i} = \frac{\sum_{j=1}^{n_e} w_{i,j}}{n_e} \quad (i = 1, \ldots, n_d)\,, \tag{8}$$

$$s_i^2 = \frac{\sum_{j=1}^{n_e} (w_{i,j} - \overline{w_i})^2}{n_e - 1} \quad (i = 1, \ldots, n_d)\,. \tag{9}$$

These two estimators are *unbiased* because they do not assume any metamodel; metamodels are only approximations so they may have important lack of fit.

Note that Dellino et al. (2009) mention that they use a crossed design, even though RSM does not require such a design. An alternative for a crossed design is the split-plot design presented by Dehlendorff et al. (2009a) or Simultaneous Perturbation Stochastic Approximation (SPSA) described by Miranda and Del Castillo (2009). Furthermore, note that the variability of the estimators is much larger for the mean than it is for the variance; e.g., under the normality assumption $var(\overline{w}) = \sigma^2/n_e$ and $var(s^2) = 2(n_e - 1)\,\sigma^4/n_e^2$; this problem is also studied by Koch et al. (1998).

In the second approach, we select a relatively small number of input combinations for the simulation model (say) n, using a space-filling design for the $k + c$ input factors (k and c still denote the number of decision and environmental factors, respectively; see again the first paragraph of Sec. 2); i.e., the environmental factors are not yet sampled from their distribution. For the larger design with N combinations, we use a space-filling design for the decision factors, but LHS for the environmental factors accounting for their distribution. We do not simulate the N combinations of this large design but we compute the Kriging predictors for the conditional means and standard deviations; i.e., in the right-hand sides of (8) and (9) we replace n_e and n_d by N_e and N_d (the large-sample analogues of the small-sample n_e and n_d) and w by $\widehat{y}$ where $\widehat{y}$ denotes the Kriging predictor computed through (4).

We shall further explain both approaches through an EOQ example in the next section. Our methodology assumes that in practice the simulation model is expensive, although we shall illustrate the two approaches through this inexpensive EOQ simulation model.

Both approaches use these estimated Kriging metamodels for the mean and standard deviation to estimate the robust optimum that minimizes the mean while satisfying a constraint on the standard deviation. Varying the value of the right-hand side for that constraint gives the Pareto frontier (see the last paragraph of Section 3 above).

This Pareto frontier is built on estimates (of the mean and standard deviation of the output). We therefore further analyze this frontier. Whereas Dellino et al. (2009) apply parametric bootstrapping, we apply *nonparametric* or *distribution-free bootstrapping*. Moreover, bootstrapping (both parametric and nonparametric) assumes that the 'original' observations are Identically and Independently Distributed (IID); see Efron and Tibshirani (1993). Because we cross the design for the decision variables and the environmental variables, the n_d observations on the output for a given combination of the environmental factors are not independent. We therefore resample the n_e *vectors* $\mathbf{w}_j$ ($j = 1, \ldots, n_e$) (with replacement, as bootstrapping requires). This resampling gives the n_e bootstrapped observations $\mathbf{w}_j^* = \left(w_{1,j}^*, \ldots, w_{n_d,j}^*\right)$; the superscript $*$ is the usual symbol for bootstrapped values. (Simar and Wilson (1998) also use distribution-free bootstrapping, albeit in the context of Data Envelopment Analysis (DEA) instead of Pareto frontiers.)

Analogous to (8) and (9) we estimate the n_d bootstrapped conditional means and variances:

$$\overline{w_i}^* = \frac{\sum_{j=1}^{n_e} w_{i,j}^*}{n_e} \quad (i = 1, \ldots, n_d)\,, \tag{10}$$

$$s_i^{2*} = \frac{\sum_{j=1}^{n_e} \left(w_{i,j}^* - \overline{w_i}^*\right)^2}{n_e - 1} \quad (i = 1, \ldots, n_d)\,. \tag{11}$$

To the estimates computed through (10) and (11) respectively we apply Kriging. The resulting two Kriging metamodels give optimum solutions for different threshold values. To reduce the sampling error in this bootstrapping, we repeat this sampling (say) B times; B is called the bootstrap sample size. This sample size gives the bootstrapped conditional averages and variances $\overline{w_{i;b}}^*$ and $s_{i;b}^{2*}$ $(b = 1, \ldots, B)$; see (10) and (11). These output data enable us to derive confidence intervals, and to account for management's risk attitude associated with the threshold value. We shall detail our procedure through an EOQ example, in the next section.

5. EOQ inventory simulation

Like Dellino et al. (2009) we apply our methodology to the simulation optimization of the EOQ inventory model. For the classic model, Zipkin (2000, pp. 30-39) uses the following symbols and assumptions: (i) The demand is known and constant, say a units per time unit. (ii) The order quantity is Q units. (iii) Total costs consist of setup cost per order, K; cost per unit purchased or produced, c; and holding cost per inventory unit per time unit, h. Management's goal is to minimize the total costs per time unit C, over an infinite time horizon.

It is easy to derive that this problem has the following *true I/O function*, which we shall use to check our simulation results:

$$C = \frac{aK}{Q} + ac + \frac{hQ}{2}\,. \tag{12}$$

This function implies that the EOQ is

$$Q_o = \sqrt{\frac{2aK}{h}}\,, \tag{13}$$

Table 1: I/O data of the classic EOQ simulation

Q	15000	22500	30000	37500	45000
C	88650.00	87641.66	87700.00	88185.00	88883.34

and the corresponding minimal cost is

$$C_o = C(Q_o) = \sqrt{2aKh} + ac\,. \tag{14}$$

In our example we use the parameter values in the classic Operations Research textbook Hillier and Lieberman (2001, pp. 936–937, 942–943): $a = 8000$, $K = 12000$, $c = 10$, and $h = 0.3$; substituting these parameter values into (13) and (14) gives $Q_o = 25298$ and $C_o = 87589$.

Following Dellino et al. (2009), we shall consider a variant of the classic EOQ model—which assumes an unknown demand rate—to tackle robustness issues. The robustness of the EOQ model is also examined by Yu (1997), who uses other criteria and other methods than we do (he uses two minmax criteria and analytical methods instead of simulation). But first we consider classic optimization.

5.1. Classic simulation optimization

In this subsection we use classic optimization; i.e., we ignore the uncertainty of the environment. Like Dellino et al. (2009) we use the following steps in our simulation experiment.

We select an experimental area for Q, namely the interval $[15000, 45000]$. Furthermore, we pick five equally spaced points in this interval, including the extreme points, 15000 and 45000. Running the simulation model with these five input values gives the total costs $C(Q_i) = C_i$ $(i = 1, \dots, 5)$; see Table 1. Based on these I/O data, we estimate the *Kriging* metamodel; see Figure 1, which also displays the true I/O function derived through (12) and the second-degree polynomial metamodel used by Dellino et al. (2009).

To validate this Kriging metamodel, we use *cross-validation*. This gives Figure 2, which shows the scatterplots for the Kriging and RSM metamodels. Because scatterplots may use scales that are misleading, Table 2 gives the relative prediction errors $\widehat{y_{-i}}/C_i$, where the subscript $-i$ means that I/O combination i is eliminated in the cross-validation, for Kriging and RSM. This validation shows that in this example Kriging does not give a better approximation than the second-order polynomial does. Our explanation uses the Taylor-series

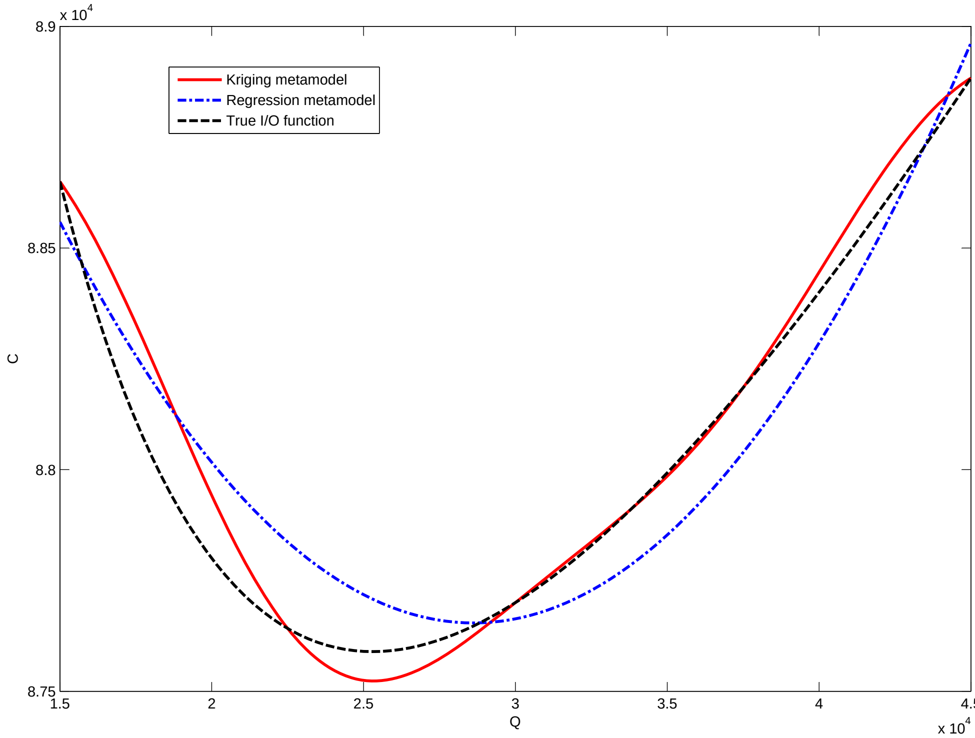


Figure 1: Kriging and RSM metamodels and the true I/O function of the classic EOQ model

argument; i.e., the EOQ model has a simple, smooth I/O function that is well approximated by a second-order polynomial in our relatively small experimental area. However, Van Beers and Kleijnen (2003) give examples in which Kriging does give better predictions than regression metamodels do. Moreover, the next paragraph will give estimated optimal order quantities and corresponding costs that are closer to the true optimal values when using Kriging instead of second-order polynomial regression. (Of course, the true optimum is known only for simple academic models such as the EOQ model; cross-validation can be applied to any metamodel.)

Table 2: Cross-validation of Kriging and RSM metamodels for EOQ cost

	Kriging		Regression	
i	$\widehat{y_{-i}}$	$\widehat{y_{-i}}/C_i$	$\widehat{y_{-i}}$	$\widehat{y_{-i}}/C_i$
1	87951.83	0.9921	87849.94	0.9910
2	88151.34	1.0058	87952.11	1.0035
3	88337.88	1.0073	87628.92	0.9992
4	88417.91	1.0026	87951.95	0.9974
5	88044.17	0.9906	87576.98	1.0078

To compute the *estimated optimum* (say) $\widehat{Q_o}$, we apply Matlab's `fmincon` to the Kriging

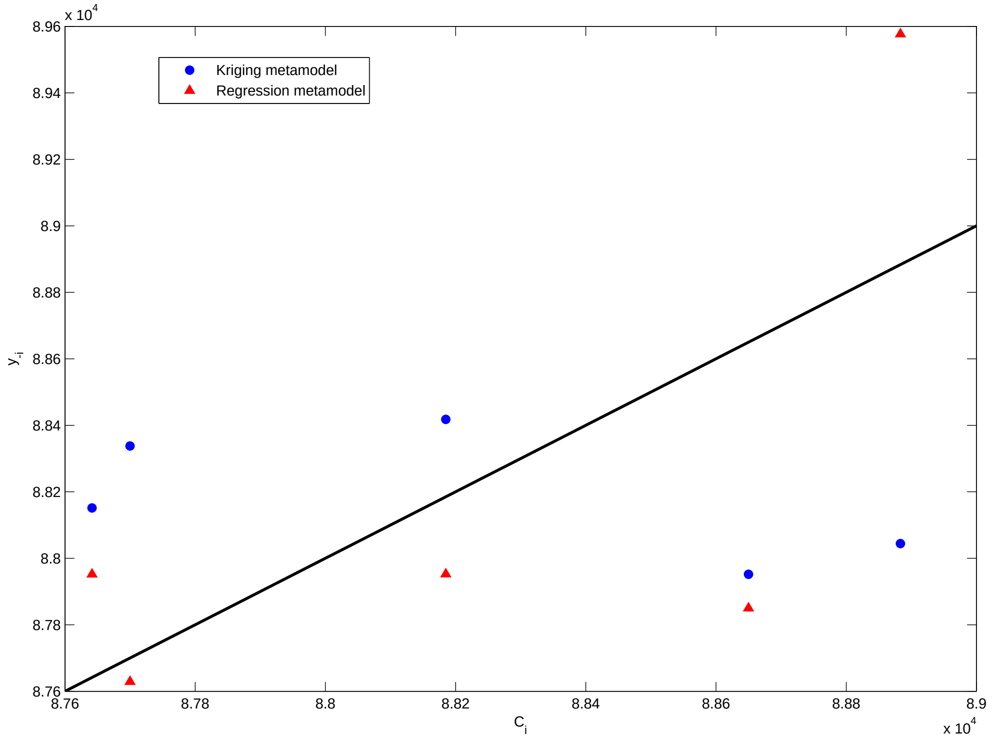


Figure 2: Scatterplots for the cross-validation of the Kriging and RSM metamodels of the classic EOQ model

metamodel for the EOQ cost (but we could have used some other solver including a global optimizer instead of a local optimizer such as `fmincon`). This gives $\widehat{Q_o} = 25337.31$ and the estimated minimal cost $\widehat{C_o} = 87523.30$. To verify these estimated optimal values, we use (13) and (14) and obtain $\widehat{Q_o}/Q_o = 25337.31/25298 = 1.0016$ and $\widehat{C_o}/C_o = 87523.30/87589 = 0.9992$; so we conclude that the estimated optimal cost and order quantity virtually equal the true optimal values. For the second-order polynomial metamodel Dellino et al. (2009) give the following results: $\widehat{Q_o} = 28636$ and $\widehat{C_o} = 887654$, so $\widehat{Q_o}/Q_o = 1.13$ and $\widehat{C_o}/C_o = 1.001$, which are slightly less accurate estimated optimal values—compared with our Kriging estimates.

Note that we also experiment with a *smaller* experimental area; i.e., a smaller Q interval. This interval gives a more accurate metamodel; the resulting estimated optimum is only 0.32% above the true EOQ and the corresponding cost virtually equals the true cost.

5.2. Robust optimization

Now we follow Dellino et al. (2009), and assume that a (demand per time unit) is an *unknown* constant; i.e., a has a Normal distribution with mean μ_a and standard deviation σ_a: $a \sim$

$N(\mu_a, \sigma_a)$. Furthermore, we assume μ_a= 8000 ('base' value used in Section 5.1), and σ_a = $0.10\mu_a$ (uncertainty about the true input parameter). This standard deviation can give negative values for a, so we resample until we get non-negative values only; this adjustment of the Normal distribution is ignored in our further analysis. We apply the two general approaches that we sketched in Section 4.2.

5.2.1. Approach 1: Kriging models for mean and standard deviation estimated from simulation I/O data

To select 'a few'—namely $n_a = 25$—values for the environmental factor a in our simulation, we use LHS ($n_a = 25$ gives 'enough' data to bootstrap later on). LHS splits the range of possible a values ($0 < a < \infty$) into n_a equally likely subranges. We use `lhsnorm` from the Matlab Statistics Toolbox to select these values from $N(\mu_a, \sigma_a)$; see The Mathworks Inc. (2005). For the control variable Q we select $n_Q = 10$ equally spaced values within $[15000, 45000]$, which is the range selected in the previous subsection. We cross these two designs for a and Q respectively, which gives 25×10 combinations of the two factors. Running the simulation model for these 250 input combinations gives an I/O table similar to Table 1. This table together with (8) and (9) gives the estimated mean and standard deviation of the cost C conditional on Q:

$$\overline{C_i} = \frac{\sum_{j=1}^{n_a} C_{i,j}}{n_a} \quad (i = 1, \ldots, n_Q\,), \tag{15}$$

$$s_i = \left[\frac{\sum_{j=1}^{n_a} \left(C_{i,j} - \overline{C_i}\right)^2}{n_a - 1} \right]^{1/2} \qquad (i = 1, \ldots, n_Q)\,. \tag{16}$$

The latter estimator is biased, because $E(\sqrt{s^2}) = E(s) \neq \sqrt{E(s^2)} = \sqrt{\sigma^2}= \sigma$; we ignore this bias.

Using (15) and (16), we fit one Kriging metamodel for the estimated mean cost—which gives $\widehat{\overline{C}}$—and one Kriging metamodel for the estimated standard deviation of cost—which gives $\widehat{s}$; obviously, each of these metamodels is based on $n_Q = 10$ observations (in the terminology of Sect. 4.1 there are 10 'old' observations). The two Kriging metamodels are shown in Figures 3 and 4, which also display the true cost function

$$E(\overline{C}) = \left(\frac{K}{Q} + c\right)\mu_a + \frac{hQ}{2}\,, \tag{17}$$

and the true standard-deviation function

$$\sigma_C = c\sigma_a + \frac{K\sigma_a}{Q}\,; \tag{18}$$

both (17) and (18) are easy to derive from (12). Figure 3 resembles the classic EOQ graph, which assumes a known demand rate. Figure 4 illustrates that the standard deviation decreases as the order quantity increases; i.e., the increased order provides a buffer against unexpected variations in the demand rate.

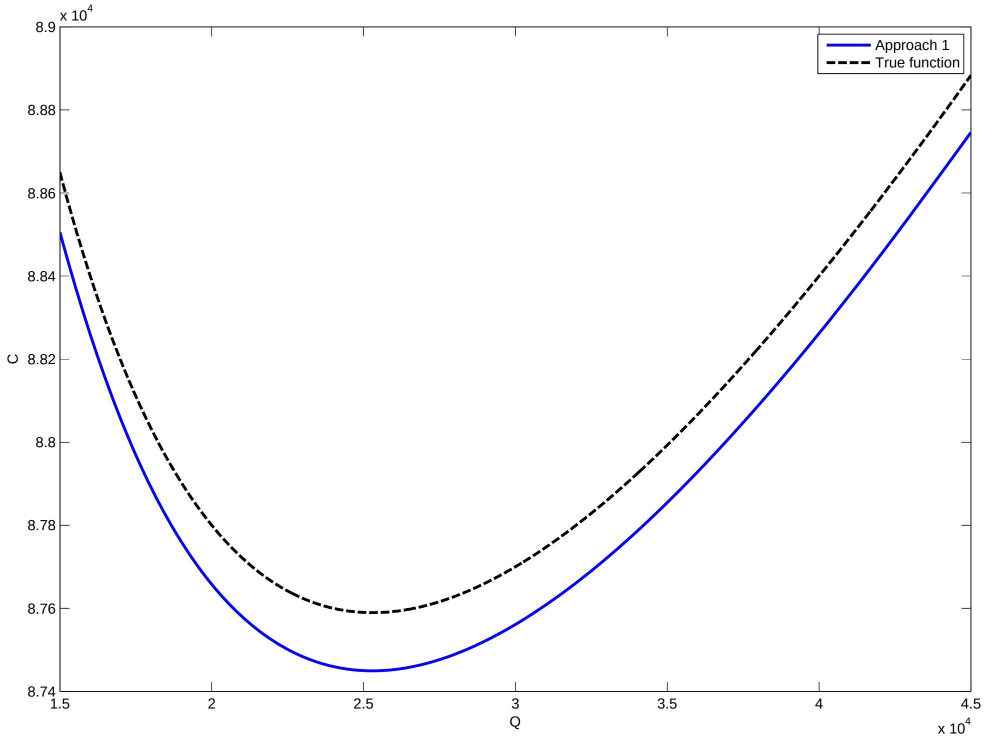


Figure 3: Kriging metamodel for mean cost in Approach 1 and the true mean cost

We validate these two metamodels—without using the true I/O functions (17) and (18)—through leave-one-out cross-validation. This validation gives the scatterplots in Figures 5 and 6, which use the symbols $\widehat{y_1}$ and $\widehat{y_2}$ for the Kriging predictors of the mean and standard deviation. Given these two figures, we accept the two Kriging metamodels.

Based on these two Kriging metamodels, we next try to find the order quantity that minimizes the mean cost, while the standard deviation does not exceed the given threshold T. We again solve this constrained optimization problem through Matlab's `fmincon`. Next we again vary this threshold, and find the set of optimal solutions that estimates the Pareto frontier; see Figure 7, which also shows the true Pareto frontier derived from (17) and (18).

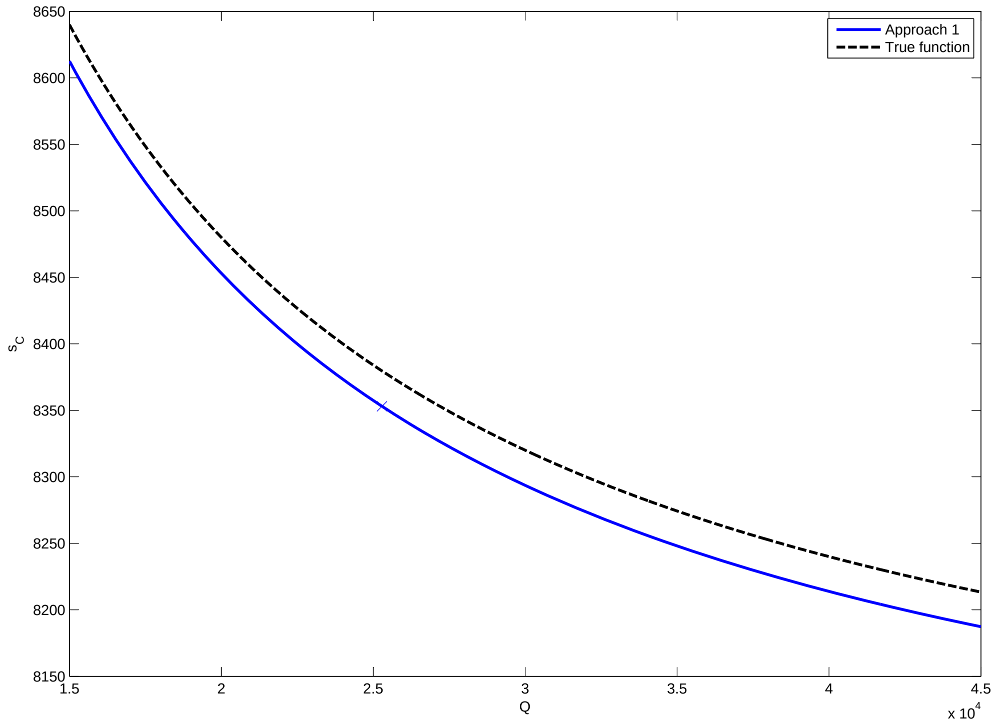


Figure 4: Kriging metamodel for the cost's standard deviation in Approach 1 and the true standard deviation

Note that we select a range for the threshold T that differs from the range in Dellino et al. (2009), because selecting the same range would have resulted in an unconstrained optimization problem so the Pareto frontier would have been a single point.

5.2.2. Approach 2: Single Kriging metamodel for mean and standard deviation

Myers and Montgomery (1995) and also Dellino et al. (2009) assume that a valid metamodel is the low-order incomplete polynomial metamodel in (1); that model implies the mean and variance displayed in (2) and (3). In a similar way we now assume that the Kriging model based on (4) gives a valid metamodel. That model has coefficients $\boldsymbol{\lambda}$ that vary with the point to be predicted; unlike (2) and (3), it does not give an explicit mean and variance.

To estimate the Kriging coefficients $\boldsymbol{\lambda}$, we select the same number of input combinations for the simulation model as we did for Approach 1; i.e., we select $n_a \times n_Q = 25 \times 10 = 250$ input combinations. To select these 250 values, we use a space-filling design in these two factors (in Approach 1 we use a space-filling design only for Q). To avoid extrapolation when using the Kriging metamodel, we select $\max a_j = \mu_a + 3\sigma_a$ and $\min a_j = \max(\mu_a - 3\sigma_a, \epsilon)$ with

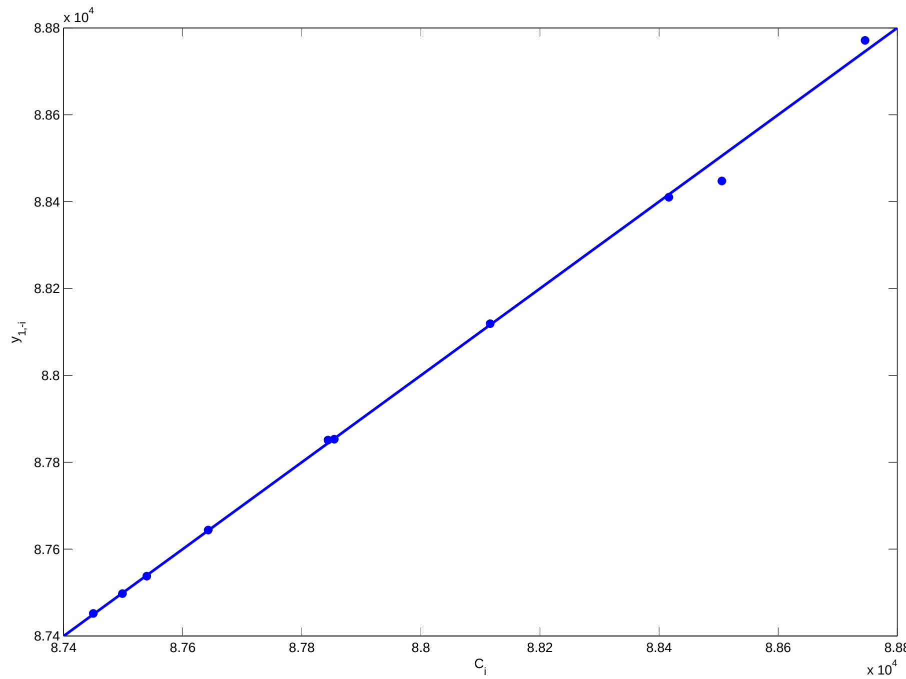


Figure 5: Scatterplot of Kriging metamodel for mean cost in Approach 1

ϵ a small positive number. After running the simulation for these 250 input combinations, the resulting I/O data give a Kriging metamodel for the costs $\widehat{C}$ as a function of the demand rate a and the order quantity Q. This metamodel is used in the following procedure:

1. Use LHS to sample $N_a \gg n_a$ values from the distribution of the environmental variable a, and use a space filling design to select $N_Q \gg n_Q$ values for the decision variable Q. We select $N_a = 100$ and $N_Q = 25$. Note that the probability of exceeding the upper bound for a is negligible; if nevertheless this event occurs, we simply take a new sample.

2. Combine the values of Step 1 into $N_a \times N_Q$ input combinations.

3. Compute the Kriging predictions $\widehat{C_{i,j}}$ $(i = 1, \ldots, N_Q\ j = 1, \ldots, N_a)$ for the combinations of Step 2, using the Kriging metamodel estimated from the smaller experiment with the simulation model with $n_a \times n_Q$ input combinations.

4. Use these predictions $\widehat{C_{i,j}}$ to estimate the N_Q conditional means and standard devia-

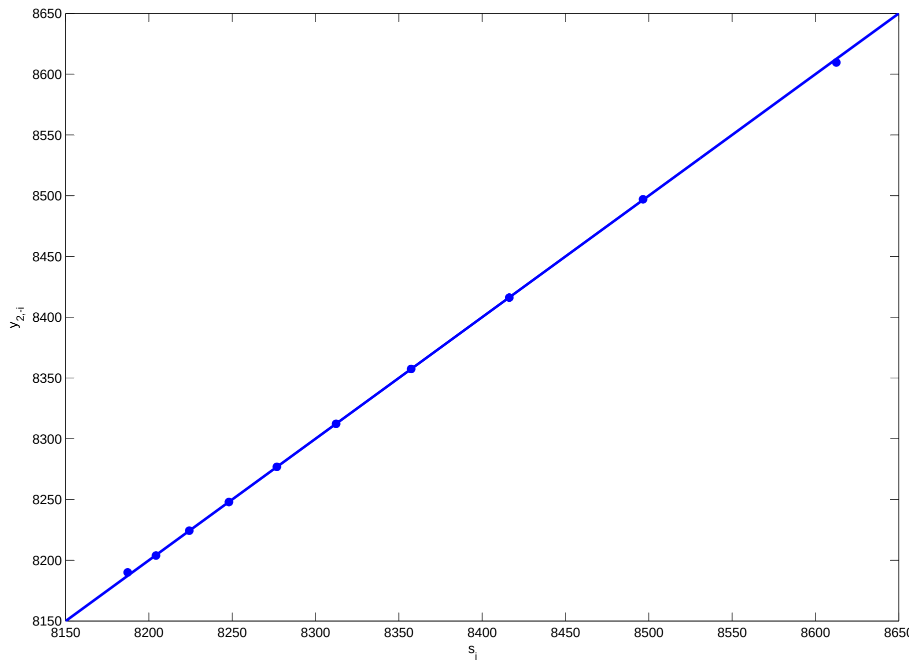


Figure 6: Scatterplot of Kriging metamodel for cost's standard deviation in Approach 1

tions of the cost C:

$$\widehat{C}_i = \frac{\sum_{j=1}^{N_a} \widehat{C_{i,j}}}{N_a} \quad (i = 1, \ldots, N_Q)\,, \tag{19}$$

$$\widehat{\sigma}_i = \frac{\sum_{j=1}^{N_a} \left(\widehat{C_{i,j}} - \widehat{C}_i\right)^2}{N_a - 1} \quad (i = 1, \ldots, N_Q)\,, \tag{20}$$

which are analogous to (15) and (16) but use a metamodel.

5. Fit one Kriging metamodel to the N_Q estimated means resulting from (19); fit another Kriging metamodel to the N_Q estimated standard deviations resulting from (20).

Figures 8 and 9 display the two Kriging models resulting from Step 5 and the true functions.

Cross-validation gives scatterplots for these two metamodels, which look very good: all points are near the 45% line, so we do not display these two figures but refer to Figures 5 and 6 (which give the scatterplots for Approach 1). Cross-validation also gives values such as

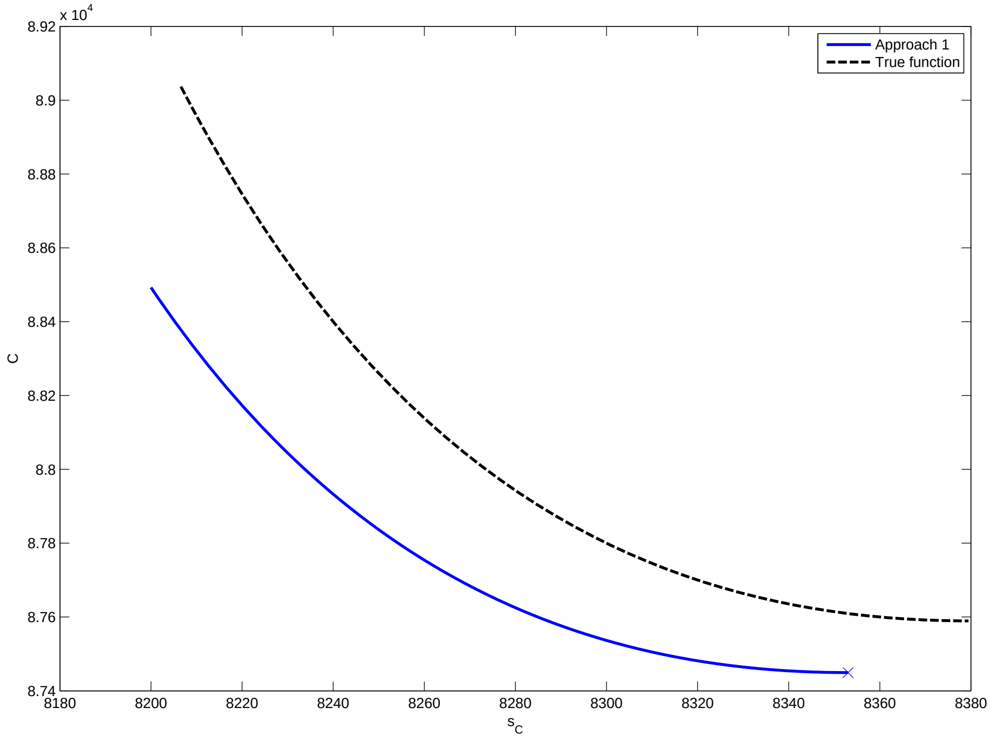


Figure 7: Estimated Pareto frontier in Approach 1 and true Pareto frontier for EOQ example

$\widehat{y_{1;(-1)}}/\widehat{C_1} = 0.9999988577$ and $\widehat{y_{2;(-25)}}/\widehat{\sigma_{25}} = 1.0000003809$, which imply very small relative prediction errors. So we accept these two Kriging metamodels as adequate approximations.

We solve the constrained optimization problem, again using `fmincon`. Next we vary the threshold T, albeit it over a range that differs from the previous range—to get interesting results, namely neither unconstrained nor infeasible results. The resulting Pareto frontier and the true frontier are displayed in Figure 10.

Finally, we compare Approaches 1 and 2. Our criteria may be their relative costs and benefits. The benefit may be the accuracy of the approach; i.e., how close is the estimated frontier to the true frontier derived from (17) and (18)? Comparing Figures 7 and 10 shows that the accuracy provided by Approach 1 is higher than that of Approach 2. The costs are the computer time needed by the two approaches: because they use two designs of the same size, their computational costs are the same. In the next subsection, however, we shall compare the two approaches through their confidence regions for the mean and standard deviation.

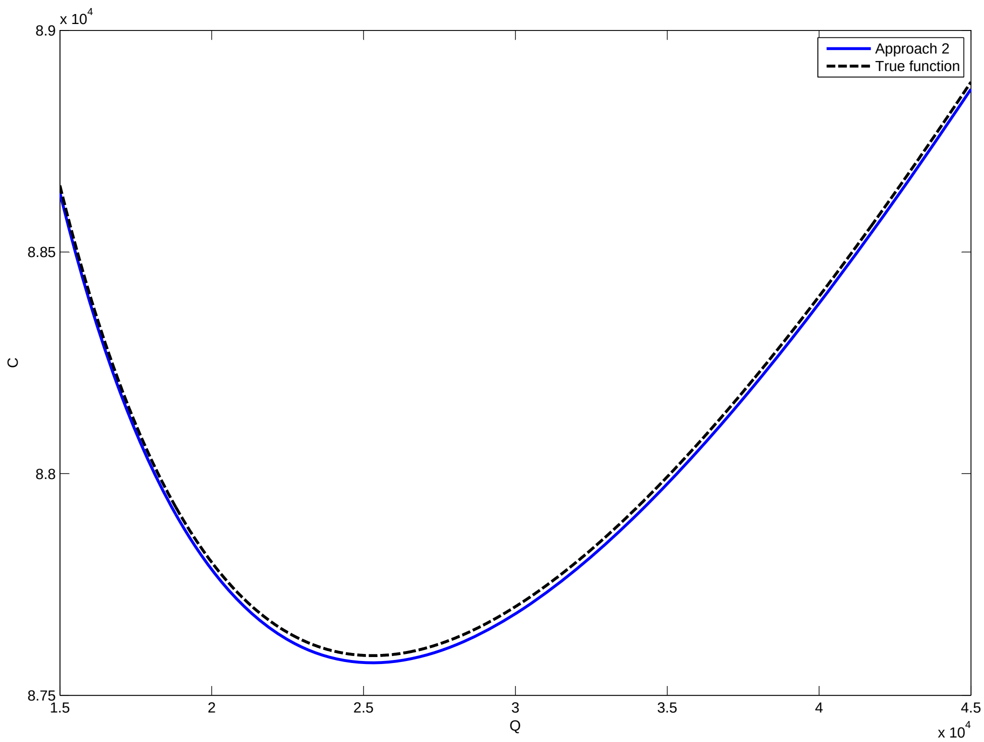


Figure 8: Kriging metamodel for mean costs based on Kriging predictions in Approach 2, and the true mean cost

5.2.3. Bootstrapped confidence region

We estimated the Pareto frontier through *random* simulation outputs $C_{i,j}$ (whether we use Approach 1 or 2), so we further analyze this frontier. As we explained through (10) and (11), we use *distribution-free bootstrapping*. This bootstrap gives $\overline{C_b}^*$ and s_b^* $(b = 1, \ldots, B)$, which gives the fitted Kriging metamodels; results for Approach 1 are given in Figure 11, where the vertical line will be explained later. This figure shows that the bootstrapped curves envelop the original curve and the true curve. Next, we use these bootstrap results as follows. (Dellino et al. (2009) use bootstrapping to derive a bundle of estimated Pareto curves, but we think that the following analysis is more relevant.)

Given the original (non-bootstrapped) Pareto frontier, management selects their preferred combination of the mean and standard deviation of the inventory cost; e.g., $\widehat{\overline{C}} = 87449.35$ and $\widehat{s} = 8353.03$, which corresponds with the 'cross' in Figure 7. Making the standard deviation not exceed its threshold implies a specific order quantity; namely, $\widehat{Q^+} = 25287.69$, which corresponds with the 'cross' in Figure 4 (displayed at the right-hand end of the estimated Pareto curve). Actually, this order quantity may give a mean and a standard deviation that

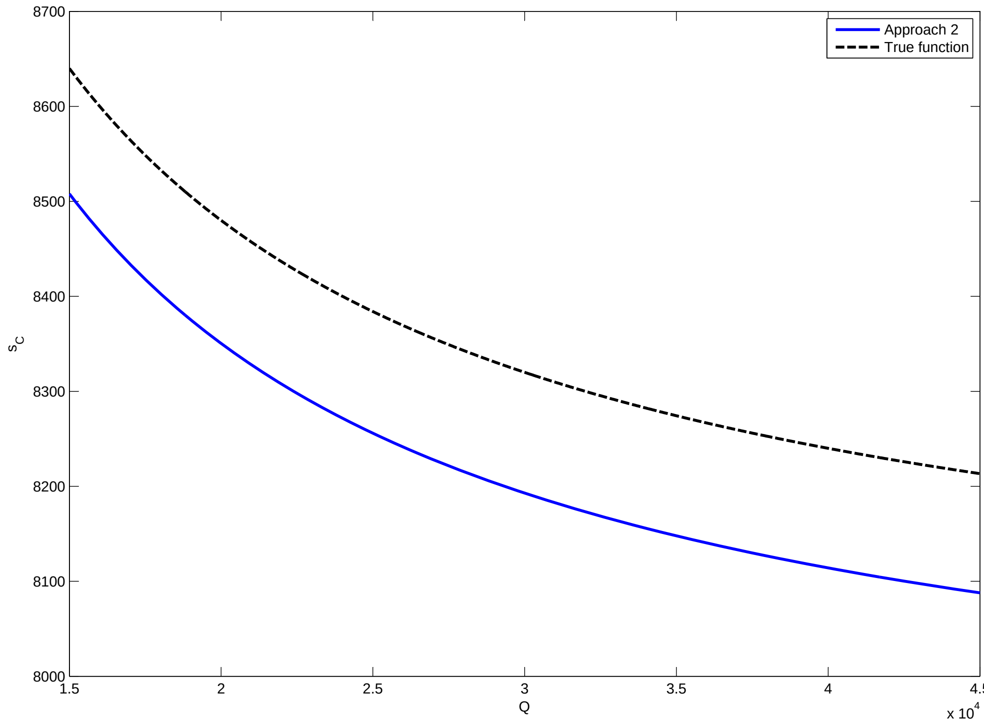


Figure 9: Kriging metamodel for the cost's standard deviation based on Kriging predictions in Approach 2, and the true function

differ from the 'original' values in Figure 11; see the vertical line in Figure 11 that is placed at $Q = \widehat{Q^{+}} = 25287.69$.

In Figure 11 this estimated Pareto-optimal order quantity $\widehat{Q^{+}}$ corresponds with B bootstrapped values for the mean and standard deviation. From these B values, we estimate a *confidence region* for the mean and standard deviation of the cost; i.e., we obtain simultaneous confidence intervals—called a confidence region—for these two outputs, as follows. We compute the distribution-free bootstrapped confidence interval (also see Efron and Tibshirani, 1993):

$$\left[\widehat{\overline{C}}^{+*}_{(\lfloor B(\alpha/2)/2 \rfloor)}, \widehat{\overline{C}}^{+*}_{(\lceil B(1-(\alpha/2))/2 \rceil)}\right]$$

where $\widehat{\overline{C}}^{+*}_{(\cdot)}$ denotes the bootstrapped average cost predicted by the Kriging metamodel in Approach 1 that corresponds with the estimated Pareto-optimal order quantity $\widehat{Q^{+}}$, the subscript ( ) denotes the order statistic (i.e., the B bootstrapped observations are ranked or sorted from smallest to largest), $\lfloor\ \rfloor$ denotes the floor function (which gives the integer part), $\lceil\ \rceil$ denotes the ceiling function (rounding upwards), $\alpha/2$ gives a two-sided confidence interval, Bonferroni's inequality implies that the type-I error rate for the interval per output

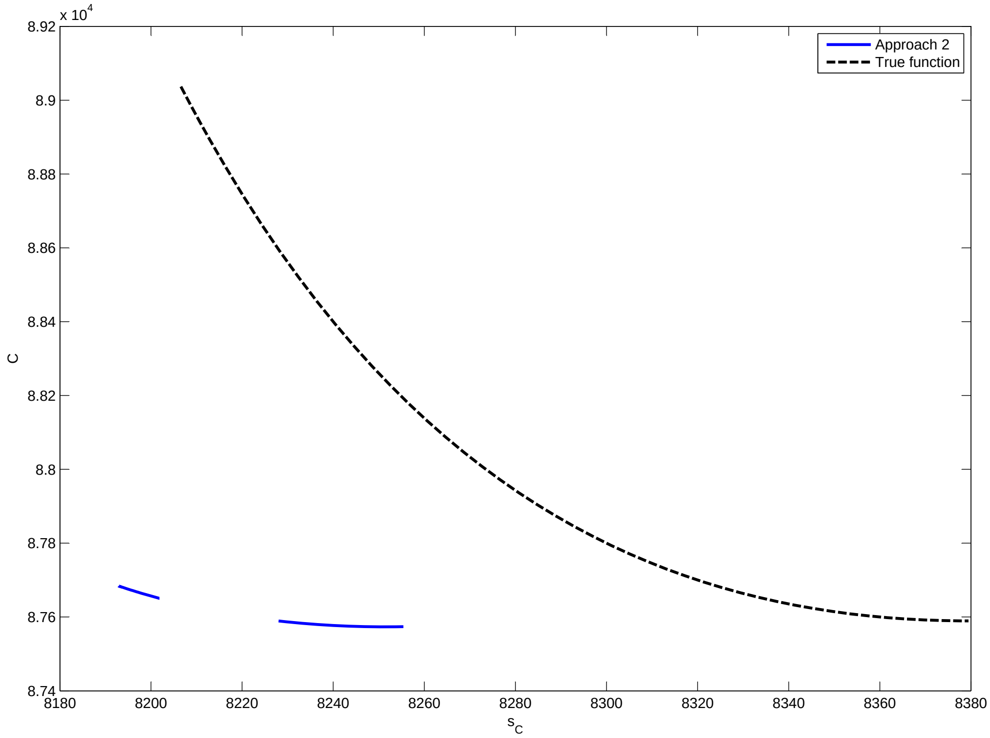


Figure 10: Estimated Pareto frontier in Approach 2 and true Pareto frontier for EOQ example

is divided by the number of outputs (which is two, namely the mean and standard deviation). Analogously we compute the following confidence interval for the standard deviation of the cost:

$$\left[\widehat{s}^{+*}_{(\lfloor B(\alpha/2)/2\rfloor)}, \widehat{s}^{+*}_{(\lceil B(1-(\alpha/2))/2\rceil)}\right] .$$

Figure 12 displays rectangular confidence regions for two points on the original estimated Pareto curve; namely, part a of the figure corresponds with the relatively small threshold value $T = 8250$ so $\widehat{Q^+} = 34940.05$ for Approach 1 and $\widehat{Q^+} = 29994.86$ for Approach 2; part b corresponds with the larger threshold value $T = 8600$ so we obtain the relatively smaller $\widehat{Q^+} = 25287.69$ for Approach 1 and $\widehat{Q^+} = 26150.43$ for Approach 2. These two threshold values reflect risk-averse and risk-seeking management. This figure shows that in this example Approach 2 gives a smaller confidence region and still covers the true point. The confidence interval for the standard deviation implies that the estimated Pareto-optimal order quantity may still give a standard deviation that violates the threshold. Confronted with this possibility, management may want to change the order quantity such that the probability of violating the threshold becomes acceptable; e.g., management may switch from the relatively small threshold in Figure 12 (part a) to the higher threshold in part b. The formalization of the problem of choosing among random outputs is beyond this article; a classic reference is Keeney and Raiffa (1976); also see the next section, which covers future research issues (see the first issue).

6. Conclusions and future research

Robust optimization of simulated systems may use *Taguchi*'s worldview, which distinguishes between decision variables that need to be optimized and environmental variables that remain uncertain but should be taken into account when optimizing. Taguchi's statistical techniques, however, may be replaced by *Kriging* metamodels (instead of low-order polynomials) and their space-filling designs such as LHS (instead of orthogonal arrays). Kriging for robust optimization may be further enhanced by *distribution-free bootstrapping*, which better enables management to make the final compromise decision. Application of this new methodology to the classic *EOQ* model shows that the classic EOQ and the robust EOQ do differ.

Future research may address the following issues. Instead of minimizing the mean under a variance constraint, we may consider alternative formulations; e.g., minimize a specific quantile of the simulation output; see Batur and Choobineh (2009), Bekki et al. (2009) and Kleijnen et al. (2009), or minimize the Conditional Value at Risk (CVaR); see Chen et al. (2009), Dehlendorff et al. (2009b) and García-González et al. (2007); the mean-variance trade-off is also criticized by Yin et al. (2009). Our Kriging metamodels may be compared with *alternative metamodels*; e.g., so-called Universal Kriging discussed in the Kriging literature, RSM (low-order polynomial linear-regression) models proposed by Dellino et al. (2009), and Generalized Linear Models proposed by Lee and Nelder (2003). In a next article we shall adjust our methodology for *random* simulation models which imply aleatory uncertainty besides epistemic uncertainty; these two types of uncertainty are discussed by De Rocquigny et al. (2008) and Helton (2009). Examples of random simulation are (s, S) models, with either explicit out-of-stock costs resulting in scalar output or a service constraint resulting in vector output (the difference $S - s$ in these models is often based on the EOQ model, discussed in this article). Finally, we hope to apply our methodology to complex *supply chain* models; also see Shukla et al. (2009) and Rao and Goldsby (2009).

Acknowledgements

Gabriella Dellino and Carlo Meloni wish to thank CentER for the financial support when they visited Tilburg University. We thank Dick den Hertog (Tilburg University) and Wim van Beers (Tilburg University) for helpful discussions on our research.

References

Ankenman, B., B. Nelson, J. Staum. 2009. Stochastic kriging for simulation metamodeling. *Oper. Res.*, accepted.

Barton, R.R., M. Meckesheimer. 2006. Metamodel-based simulation optimization. S.G. Henderson, B.L. Nelson, eds. *Handbook in OR & MS* **13** Elsevier. 535–574.

Batur, D., F. Choobineh. 2009. A quantile-based approach to system selection. *Eur. J. of Oper. Res.*, in press.

Bekki, J.M., J.W. Fowler, G.T. Mackulak, M. Kulahci. 2009. Simulation-based cycle-time quantile estimation in manufacturing settings employing non-FIFO dispatching policies. *J. of Simulation* **3** 69–83.

Beyer, H., B. Sendhoff. 2007. Robust optimization—a comprehensive survey. *Comput. Methods in Appl. Mech. and Engrg.* **196** 3190–3218.

Chen, Y., M. Xu, Z.G. Zhang. 2009. A risk-averse newsvendor model under CVaR criterion. *Oper. Res.* **57** 1040–1044.

Cressie, N.A.C. 1993. *Statistics for spatial data* (revised ed.). John Wiley & Sons, New York.

Dehlendorff, C., M. Kulahci, K. Andersen. 2009. Designing simulation experiments with controllable and uncontrollable factors for applications in health care. *J. of the Roy. Statist. Soc.*, under review.

Dehlendorff, C., M. Kulahci, S. Merser, K. Andersen. 2009. Conditional value at risk as a measure for waiting time in simulations of hospital units. Working paper, DTU Informatics, Technical University of Denmark.

Del Castillo, E. 2007. *Process optimization: a statistical approach.* Springer, New York.

Dellino, G., J.P.C. Kleijnen, C. Meloni. 2009. Robust optimization in simulation: Taguchi and Response Surface Methodology. *CentER Discussion Paper*, Tilburg University, Tilburg, the Netherlands. http://center.uvt.nl/staff/kleijnen/papers.html.

De Rocquigny, E., N. Devictor, S. Tarantola. 2008. Uncertainty settings and natures of uncertainty. E. De Rocquigny, N. Devictor and S. Tarantola, eds. *Uncertainty in industrial practice*. Wiley, Chichester, England.

Efron, B., R.J. Tibshirani. 1993. *An introduction to the bootstrap*. Chapman & Hall, New York.

García-González, J., E. Parrilla, A. Mateo. 2007. Risk-averse profit-based optimal scheduling of a hydro-chain in the day-ahead electricity market. *Eur. J. of Oper. Res.* **181** 1354–1369.

Helton, J.C. 2009. Conceptual and computational basis for the quantification of margins and uncertainty. SANDIA Report SAND2009-3055.

Hillier, F.S., G.J. Lieberman. 2001. *Introduction to Operations Research; seventh edition*. McGraw Hill, Boston.

Keeney, R.L., H. Raiffa. 1976. *Decisions with multiple objectives: preferences and value tradeoffs*. John Wiley & Sons, Inc., New York.

Kleijnen, J.P.C. 2008. *Design and analysis of simulation experiments*. Springer, New York.

Kleijnen, J.P.C. 2009. Sensitivity analysis of simulation models. J.J. Cochran, L.A. Cox, P. Keskinocak, J.P. Kharoufeh, J.C. Smith, eds. *Wiley Encyclopedia of Operations Research and Management Science*. Wiley, New York.

Kleijnen, J.P.C., H. Pierreval, J. Zhang. 2009. Methodology for determining the acceptability of given designs in uncertain environments. Working paper, Tilburg University. http://center.uvt.nl/staff/kleijnen/papers.html.

Kleijnen, J.P.C., S.M. Sanchez, T.W. Lucas, T.M. Cioppa. 2005. State-of-the-art review: a user's guide to the brave new world of designing simulation experiments. *INFORMS J. Computing* **17** 263–289.

Koch, P.K., D. Mavris, J.K. Allen, F. Mistree. 1998. Modeling noise in approximation-based robust design: a comparison and critical discussion. *ASME Design Engineering Technical Conferences, Atlanta, September 1998.*

Lee, Y., J.A. Nelder. 2003. Robust design. *J. of Quality Tech.* **35** 2–12.

Lee, K.H., G.J. Park. 2006. A global robust optimization using Kriging based approximation model. *J. of the Japan Soc. of Mech. Engrg.* **49** 779–788.

Lehman, J.S., T.J. Santner, W.I. Notz. 2004. Designing computer experiments to determine robust control variables. *Statistica Sinica* **14** 571–590.

Lophaven, S.N., H.B. Nielsen, J. Søndergaard. 2002. DACE: a MATLAB Kriging toolbox (version 2.0). Technical Report IMM-TR-2002-12, Technical University of Denmark, Lyngby.

Miranda, A.K., E. del Castillo. 2009. Robust parameter design optimization of simulation experiments using stochastic perturbation methods. *J. of the Oper. Res. Soc.* (to appear).

Myers, R.H., D.C. Montgomery. 1995. *Response surface methodology: process and product optimization using designed experiments.* John Wiley & Sons, New York.

Nair, V.N., editor. 1992. Taguchi's parameter design: a panel discussion. *Technometrics* **34** 127–161.

Oberguggenberger, M., J. King, B. Schmelzer. 2009. Classical and imprecise probability methods for sensitivity analysis in engineering: a case study. *Internat. J. of Approximate Reasoning* **50** 680–693.

Park, G-J., T-H. Lee, K.H. Lee, K-H. Hwang. 2006. Robust design: an overview. *AIAA J.* **44** 181–191.

Rao, S., T.J. Goldsby. 2009. Supply chain risks: a review and typology. *Internat. J. of Logist. Management* **20** 97–123.

Sacks, J., W.J. Welch, T.J. Mitchell, H.P. Wynn. 1989. Design and analysis of computer experiments. *Statist. Sci.* **4** 409–435.

Santner, T.J., B.J. Williams, W.I. Notz. 2003. *The design and analysis of computer experiments*. Springer-Verlag, New York.

Shukla, S.K., M.K. Tiwari, H.-D. Wan, R. Shankar. 2009. Optimization of the supply chain network: simulation, Taguchi, and psychoclonal algorithm embedded approach. *Computers & Industrial Engineering* (in press).

Simar, L., P.W. Wilson. 1998. Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models. *Management Science* **44** 49–61.

Stinstra, E., D. den Hertog. 2008. Robust optimization using computer experiments. *Eur. J. of Oper. Res.* **191** 816–837.

Taguchi, G. 1987. *System of experimental designs, volumes 1 and 2*. UNIPUB/ Krauss International, White Plains, New York.

The MathWorks Inc. 2005. *Statistics Toolbox*. Natick, Massachusetts.

van Beers, W.C.M., J.P.C. Kleijnen. 2003. Kriging for interpolation in random simulation. *J. of the Oper. Res. Soc.* **54** 255–262.

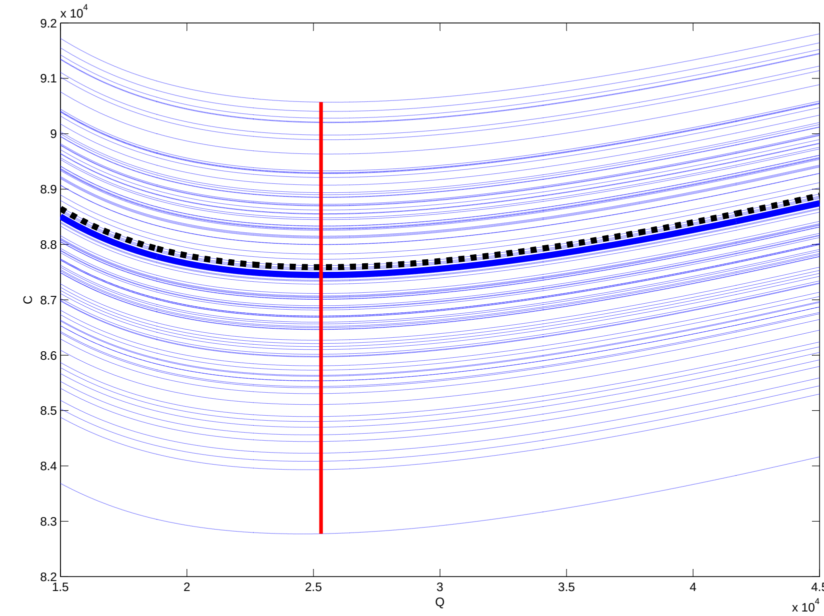
Wu, C.F.J., M. Hamada. 2000. *Experiments; planning, analysis, and parameter design optimization*. Wiley, New York.

Wu. J., J. Li, S. Wang, T.C.E. Cheng. 2008. A note on mean-variance analysis of the newsvendor model with stockout cost *Omega*, in press.

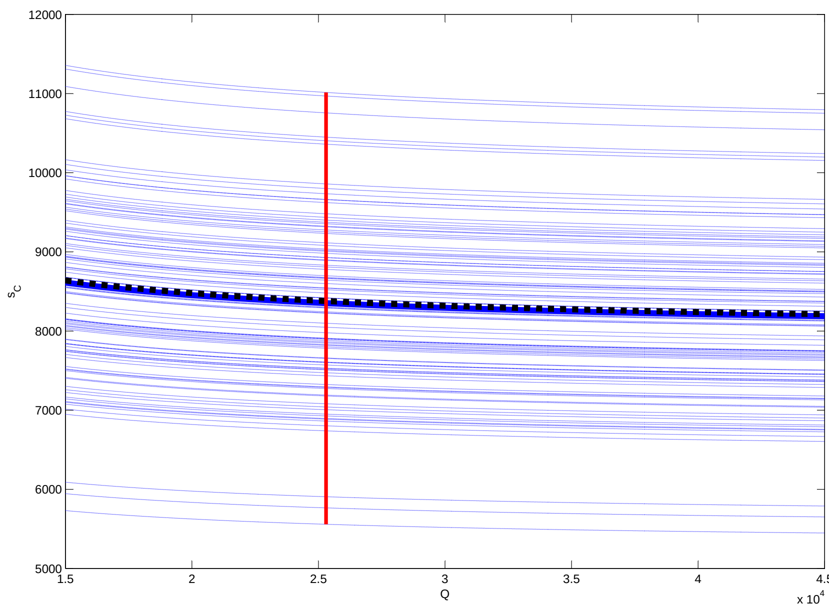
Yin, Y., S.M. Madanat, X.-Y. Lu. 2009. Robust improvement schemes for road networks under demand uncertainty. *Eur. J. of Oper. Res.* **198** 470–479.

Yu, G. 1997. Robust economic order quantity models. *Eur. J. of Oper. Res.* **100** 482–493.

Zipkin, P.H. 2000. *Foundations of inventory management*. McGraw-Hill.

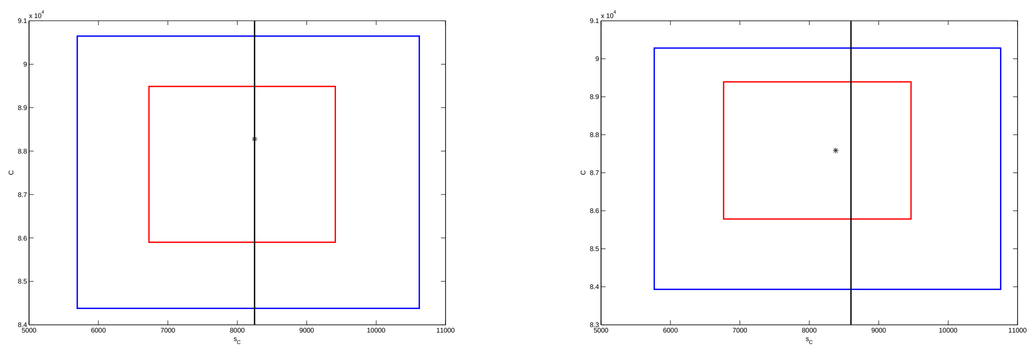


(a)



(b)

Figure 11: (a) Bootstrapped estimated costs, original metamodel in Approach 1 (heavy curve) and true cost (dotted curve); (b) Bootstrapped estimated standard deviations of cost, original metamodel in Approach 1 (heavy curve) and true standard deviation (dotted curve)



(a) (b)

Figure 12: Confidence regions for σ_C (on x-axis) and $E(\overline{C})$ (on y-axis) based on bootstrapped Kriging in Approach 1 (outer rectangle) and Approach 2 (inner rectangle) at (a) $T = 8250$ and (b) $T = 8600$ (see the vertical line). '$*$' denotes the 'true' solution based on (17) and (18)